```
=> d his

     (FILE 'HOME' ENTERED AT 15:46:37 ON 10 MAY 2005)
     FILE 'CA' ENTERED AT 15:46:47 ON 10 MAY 2005
L1 318345 S (PEAK OR INTENSITY OR LINE OR MASS OR SIGNAL)(5A)(DATA OR
          INFORMATION OR POSITION OR VALUE OR SPECTRA# OR SPECTRUM)
L2 138514 S (PEAK OR INTENSITY OR LINE OR MASS OR SIGNAL)(5A)(IDENTIF? OR
          SEARCH? OR MINE# OR MINING OR COMPAR? OR FIT? OR INDEX? OR SCORE#
          OR SCORING OR SCREEN? OR EXAMIN?)
L3 412000 S (DATA OR INFORAMTION OR POSITION OR VALUE OR SPECTRA# OR
          SPECTRUM)(5A)(IDENTIF? OR SEARCH? OR MINE# OR MINING OR COMPAR? OR
          FIT? OR INDEX? OR SCORE# OR SCORING OR SCREEN? OR EXAMIN?)
L4 112356 S L1-3 AND MASS SPECTR?
L5  17200 S L4 AND(DATABASE OR DATA BASE OR COMPIL? OR LIBRARY OR GROUP)
L6   1286 S L5 AND(AUTOMAT? OR COMPUTER OR MICROPROCESSOR OR ALGORITHM)
L7    786 S L6 NOT PY>2000
L8     14 S L6 NOT L7 AND PATENT/DT AND PY<2001
L9   6643 S (PEAK OR INTENSITY OR LINE OR MASS OR SIGNAL)(5A)(MATCH? OR
          CLASSIF?)
L10 13416 S (DATA OR INFORAMTION OR POSITION OR VALUE OR SPECTRA# OR
          SPECTRUM)(5A)(MATCH? OR CLASSIF?)
L11   348 S L9-10 NOT L1-3 AND MASS SPECTR?
L12    79 S L11 AND(DATABASE OR DATA BASE OR COMPIL? OR LIBRARY OR GROUP)
L13    61 S L11 AND(AUTOMAT? OR COMPUTER OR MICROPROCESSOR OR ALGORITHM)
L14    67 S L12-13 NOT PY>2000
L15     0 S L12-13 NOT L14 AND PATENT/DT AND PY<2001
L16   867 S L7-8,L14
L17   348 S L16 AND(SIMPLE MATCHING OR AUTOCLASSIF? OR COMPUTER ASSISTED OR
          SUBSTRUCTURE OR NEUTRAL LOSS OR EXPERT OR INTERPRET? OR BINARY
          ENCOD? OR PHEROMONE OR DE NOVO OR DAUGHTER OR RANK? OR UNKNOWN OR
          PATTERN RECOGNI? OR STRUCTURE ELUCID? OR FORWARD REVERSE OR
          STRUCTURAL ANALOG OR PETROLEUM)
L18    38 S L16 AND(PROBABILITY OR CONSTRAINED OR MACPROMASS)
L19   508 S L16 NOT L17-18
L20    26 S L19 AND(POST TRANSLATIONAL OR ANALOG SPECTRA OR MUTATION TOLERANT
          OR UNEXPECTED OR MINING OR COMPUTER MATCHING)
L21     2 S L19 AND (SCREENING PEPTIDE OR QMASS)
L22   387 S L17-18,L20-21

=> d bib,ab 1-387 l22

L22  ANSWER 15 OF 387  CA  COPYRIGHT 2005 ACS on STN
AN   133:185290  CA
TI   Storing spectral data from mass spectrometers
IN   Franzen, Jochen
PA   Bruker Daltonik G.m.b.H., Germany
SO   Brit. UK Pat. Appl., 11 pp.
PI   GB 2342498          A1    20000412    GB 1999-23561         19991005
     US 6624408          B1    20030923    US 1999-407729        19990928
PRAI DE 1998-19845729     A     19981005
AB   The invention consists of combining all or selected daughter and
     granddaughter spectra of a parent ion in an ion trap over several
     generations in one combined descendants spectrum.  This combined
     descendants spectrum can be depicted as a graphic or a list.  The refs.
```

to origin can be plotted on the combined descendants spectrum. For biopolymers, where the loss of fragments can be **identified** due to their **mass**, the names or abbreviations of lost mol. fragments can be entered. The criteria for selection of the spectra can be predefined; in this way, the spectra can be depicted and even scanned **automatically**. The combined descendants **spectrum** facilitates **comparison** with **spectrum library data** from other types of **mass spectrometer**, e.g. tandem in space **mass spectrometers**, to enable **identification** or structural elucidation of the parent ion.

L22 ANSWER 35 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 132:90254 CA
TI Locating and identifying posttranslational modifications by in-source decay during MALDI-TOF **mass spectrometry**
AU Lennon, John J.; Walsh, Kenneth A.
CS Department of Biochemistry, University of Washington, Seattle, WA, 98195-7350, USA
SO Protein Science (1999), 8(11), 2487-2493
AB A technique is described for identifying and locating posttranslational modifications (PTMs) in peptides and proteins of known sequence by **interpretation** of cn ion signals generated by in-source decay during delayed ion extn. in matrix-assisted laser desorption/ionization time-of-flight **mass spectrometry**. Sites of phosphorylation in seven synthetic peptides were detd., as was the location of both the heme **group** and N,N,N-trimethyllysine in yeast cytochrome c. A semi-**automated data** anal. process facilitates the **identification** of segments of the sequence on each side of the PTM, permitting its placement at the junction of the segments and definition of the added mass. A graphical display facilitates illustration of both the location and mass of the PTM.

L22 ANSWER 45 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 131:195866 CA
TI High-throughput **mass spectrometric** discovery of protein **post-translational** modifications
AU Wilkins, Marc R.; Gasteiger, Elisabeth; Gooley, Andrew A.; Herbert, Ben R.; Molloy, Mark P.; Binz, Pierre-Alain; Ou, Keli; Sanchez, Jean-Charles; Bairoch, Amos; Williams, Keith L.; Hochstrasser, Denis F.
CS Macquarie University Centre for Analytical Biotechnology and Australian Proteome Analysis Facility, Macquarie University, Sydney, NSW 2109, Australia
SO Journal of Molecular Biology (1999), 289(3), 645-657
AB The availability of genome sequences, affordable **mass spectrometers** and high-resoln. two-dimensional gels has made possible the identification of hundreds of proteins from many organisms by peptide mass fingerprinting. However, little attention has been paid to how information generated by these means can be utilized for detailed protein characterization. Here we present an approach for the systematic characterization of proteins using **mass spectrometry** and a software tool Find Mod. This tool **examines** peptide **mass** fingerprinting **data** for **mass** differences between empirical and theor. peptides. Where mass differences correspond to a **post-translational** modification, intelligent rules are applied to predict the amino acids in the peptide,

if any, that might carry the modification. Find Mod rules were constructed by examg. 5153 incidences of **post-translational** modifications documented in the SWISS-PROT **database**, and for the 22 **post-translational** modifications currently considered (acetylation, amidation, biotinylation, C-mannosylation, deamidation, flavinylation, farnesylation, formylation, geranyl-geranylation, gamma-carboxyglutamic acids, hydroxylation, lipoylation, methylation, myristoylation, N -acyl diglyceride (tripalmitate), O-GlcNAc, palmitoylation, phosphorylation, pyridoxal phosphate, phospho-pantetheine, pyrrolidone carboxylic acid, sulfation) a total of 29 different rules were made. These consider which amino acids can carry a modification, whether the modification occurs on N-terminal, C-terminal or internal amino acids, and the type of organisms on which the modification can be found. We illustrate the utility of the approach with proteins from 2-D gels of Escherichia coli and sheep wool, where **post-translational** modifications predicted by Find Mod were confirmed by MALDI post-source decay peptide fragmentation. As the approach is amenable to **automation**, it presents a potentially large-scale means of protein characterization in proteome projects.

L22 ANSWER 62 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 129:105694 CA
TI The identification of peptide modifications derived from gel-separated proteins using electrospray triple quadrupole and ion trap analyses
AU Swiderek, Kristine M.; Davis, Michael T.; Lee, Terry D.
CS Beckman Research Inst., City Hope, Duarte, CA, USA
SO Electrophoresis (1998), 19(6), 989-997
AB Microspray tandem **mass spectrometry** (MS/MS) in combination with **database** search routines has become a powerful tool for the identification of proteins from femtomole amts. of material following gel electrophoresis and in-gel digestion procedures. Artifactual modification of susceptible residues can arise during gel electrophoresis, leading to unexpected peptide mass shifts during mass anal. Collision-induced dissocn. (CID) spectra generated from these derivatized peptides can defy direct **interpretation** by **automated database** search routines and remain unidentified. The authors evaluate the MS/MS spectra of peptides carrying oxidized derivs. of Trp and Met residues, and various modifications of Cys. The authors demonstrate that certain of these modifications generate characteristic fragmentation patterns or "fingerprints", during CID anal., the knowledge of which can facilitate the **interpretation** of the spectra. The authors show that these signature fragment ions are predominantly produced during the CID anal. of singly charged ions although they can be obsd. in the MS/MS spectra of the doubly charged species as well. In other cases, the CID spectrum lacks a characteristic fingerprint and the modification remains silent. CID spectra of related peptides, differing only by their modifications, are similar and all or part of the fragment ion spectra will have shifted by a discreet **mass**, which facilitates the **identification** of the modified residue. At the same time, the **comparison** of related **spectra** can prevent misinterpretations such as the assignment of a residue mass to the wrong amino acid or a **neutral loss** fragment ion to a y- or b-ion.

L22 ANSWER 69 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 128:250175 CA

TI New **Computer** Aided Methods for Revealing Structural Features of **Unknown** Compounds Using Low Resolution **Mass Spectra**
AU Lebedev, Konstantin S.; Cabrol-Bass, Daniel
CS Institute of Organic Chemistry, Siberian Branch of Russian Academy of Science, Novosibirsk, 630090, Russia
SO Journal of Chemical Information and Computer Sciences (1998), 38(3), 410-419
AB Two new **computer** methods designed to reveal structural features of **unknown** compds. by low resoln. **mass spectra** are presented. Both methods use the results of a **spectral** similarity **search** in a **mass spectral database**. The 1st one proceeds by intersecting selected structures to find maximal common **substructures**, while the 2nd proceeds by decompg. these structures to derive fragments following a model of primary fragmentation of org. mols. Reliability of the revealed fragments is estd. by **comparing** an **unknown** compd.'s **spectrum** with the computed spectral images of each fragment. The usefulness and limitations of the two proposed methods are estd. by using a set of test examples. In many cases the two methods are complementary, whereas overall, the 2nd looks more promising both for revealing large structural fragments and for generation of candidate structures, because the fragments revealed have only one or two free valences and rarely overlap one another.

L22 ANSWER 80 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 127:132915 CA
TI Sequence **database** searches via **de novo** peptide sequencing by tandem **mass spectrometry**
AU Taylor, J. Alex; Johnson, Richard S.
CS Dep. Biochem., Univ. Washington, Seattle, WA, 98195-7350, USA
SO Rapid Communications in Mass Spectrometry (1997), 11(9), 1067-1075
AB A method is described for **searching** protein sequence **databases** using tandem **mass spectra** of tryptic peptides. The approach uses a **de novo** sequencing **algorithm** to derive a short list of possible sequence candidates which serve as query sequences in a subsequent homol.-based **database** search routine. The sequencing **algorithm** employs a graph theory approach similar to previously described sequencing programs. In addn., amino acid compn., peptide sequence tags, and incomplete or ambiguous Edman sequence data can be used to aid in the sequence detns. Although sequencing of peptides from tandem **mass spectra** is possible, one of the frequently encountered difficulties is that several alternative sequences can be deduced from one spectrum. Most of the alternative sequences, however, are sufficiently similar for a homol.-based sequence **database** search to be possible. Unfortunately, the available protein sequence **database** search **algorithms** (e.g. Blast or FASTA) require a single unambiguous sequence as input. Here we describe how the publicly available FASTA **computer** program was modified in order to search protein **databases** more effectively in spite of the ambiguities intrinsic in **de novo** peptide sequencing **algorithms**.

L22 ANSWER 99 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 124:305986 CA
TI **Automatic** recognition of substance classes from **data** obtained by gas chromatography/**mass spectrometry**

AU Varmuza, K.; Stancl, F.; Lohninger, H.; Werther, W.
CS Dep. Chemometrics, Technical Univ. Vienna, Vienna, A-1060, Austria
SO Laboratory Automation and Information Management (1996), 31(3), 225-30
AB The combination of gas chromatog. and **mass spectrometry** is one of the most powerful instrumental techniques for analyses of complex samples. A bottleneck is the **interpretation** of the huge amt. of data produced during an anal. A new software program, MSclass, contains classifiers for the **automatic** recognition of ~80 chem. **substructures** or classes of compds. directly from low resoln. **mass spectra**.

L22 ANSWER 113 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 123:270262 CA
TI **Mass spectra interpretation** system including **spectra** extraction
IN Gray, Zachary A.; Abel, Roger H.
PA Hewlett Packard Company, USA
SO U.S., 19 pp.
PI US 5453613 A 19950926 US 1994-327166 19941021
PRAI US 1994-327166 A 19941021
AB **Mass spectral** analyzer systems and a method for providing **automated** discovery, deconvolution, and **identification** of **mass spectra** are described. Conventionally acquired **mass data** files are re-sorted from chronol. to primarily ion-mass order and secondarily to chronol. order within each ion-mass grouping. For each ion-mass measured, local **peaks** or max. are **identified** through an integrator means. All local max. are then sorted and partitioned such that a set of deconvoluted spectra is obtained such that each element of the set constitutes an identifiable compd. Compds. may then be matched to ref. spectra in **library** datafiles by conventional probabilistic matching routines.

L22 ANSWER 116 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 123:187353 CA
TI Chemical **substructure identification** by **mass spectral library searching**
AU Stein, Stephen E.
CS NIST Mass Spectrometry Data Cent., Gaithersburg, MD, USA
SO Journal of the American Society for Mass Spectrometry (1995), 6(8), 644-55
AB A **library**-search procedure that identifies structural features of an **unknown** compd. from its electron-ionization **mass spectrum** is described. Like other methods, this procedure 1st retrieves **library** compds. whose spectra are most similar to the spectrum of an **unknown** compd. If then deduces structural features of the **unknown** compd. from the chem. structures of the retrievals. Unlike other methods, the significance of each retrieved spectrum is weighted according to its similarity to the spectrum of the **unknown** compd. Also, a **peaks**-in-common **screening** step serves to reduce search times and an optimized dot product function provides the match factor. If the mol. wt. of the **unknown** compd. is provided, the identification of certain **substructures** can be improved by including **neutral loss** peaks. Correlations between the presence of a **substructure** in a test searching the NIST/EPA/NIH ref. **library** with a 7891. compd. test set. These correlations allow the estn. of **probabilities** of **substructure** occurrence and absence in an **unknown** compd. from the results of a **library** search. This method may be viewed

as an optimization of the K-nearest neighbor method of Isenhour and co-workers, with improvements that arise from **spectrum screening, peak** scaling, an optimal distance measure, a relative-distance weighting scheme, and a larger ref. **library.**

L22 ANSWER 118 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 123:48602 CA
TI Error-Tolerant Identification of Peptides in Sequence **Databases** by Peptide Sequence Tags
AU Mann, M.; Wilm, M.
CS Protein Peptide Group, European Molecular Biology Laboratory, Heidelberg, D-69012, Germany
SO Analytical Chemistry (1994), 66(24), 4390-9
AB The authors demonstrate a new approach to the **identification** of **mass spectrometrically** fragmented peptides. A fragmentation **spectrum** usually contains a short, easily **identifiable** series of sequence ions, which yields a partial sequence. This partial sequence divides the peptide into three parts-regions 1, 2, and 3-characterized by the added mass m1 of region 1, the partial sequence of region 2, and the added mass m3 of region 3. The authors call the construct, m1 partial sequence m3, a "peptide sequence tag" and show that it is a highly specific identifier of the peptide. An **algorithm** developed here that uses the sequence tag to find the peptide in a sequence **database** is up to 1 million-fold more discriminating than the partial sequence information alone. Peptides can be identified even in the presence of an **unknown** post-translational modification or an amino acid substitution between an entry in the sequence **database** and the measured peptide. These concepts are demonstrated with model and practical examples of electro-spray **mass spectrometry/mass spectrometry** of tryptic peptides. Just two to three amino acid residues derived by fragmentation are enough to identify these peptides. In peptide mapping applications, even less information is necessary.

L22 ANSWER 138 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 120:318607 CA
TI Identification of proteins in polyacrylamide gels by **mass spectrometric** peptide mapping combined with **database** search
AU Mortz, Ejvind; Vorm, Ole; Mann, Matthias; Roepstorff, Peter
CS Dep. Mol. Biol., Odense Univ., Odense, 5230, Den.
SO Biological Mass Spectrometry (1994), 23(5), 249-61
AB **Mass spectrometric** peptide mapping of proteins sepd. by one-dimensional SDS-PAGE has been investigated. The best results are obtained after blotting of the proteins onto polyvinylidene difluoride membranes followed by enzymic digestion of the protein on the membrane. The peptide maps were investigated in terms of completeness and applicability for protein identification using a previously developed **database** search program as well as for the possibility for full characterization of covalent modifications in the proteins. The most complete peptide maps were obtained when the proteins were reduced and alkylated on the membrane prior to enzymic digestion followed by sepn. of the resulting mixt. by HPLC prior to **mass spectrometric** anal. Such peptide maps cover up to 98% of the sequence and consequently may allow complete characterization of **post-translational** modifications in

proteins for which the amino acid sequence is known. The fastest and most sensitive procedure to obtain peptide maps sufficient for protein identification was direct anal. of the extd. peptide mixt. by matrix-assisted laser desorption ionization (MALDI) **mass spectrometry**. The use of external and internal calibration of MALDI **spectra** for **database searches** is evaluated as well as the possibility of including a post-calibration routine within the search program.

L22 ANSWER 144 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 119:48841 CA
TI **Computer**-aided **interpretation** of **mass spectra** using **databases** with **spectra** and structures. I. Structure **searches**
AU Henneberg, D.; Weimann, B.; Zalfen, U.
CS Max-Planck-Inst. Kohlenforsch., Muelheim ander Ruhr, Germany
SO Organic Mass Spectrometry (1993), 28(3), 198-206
AB For **databases** contg. spectra and structures of the ref. compds., structural descriptors (fragments) are derived that are used for structure searches in the **databases**. The 190 fragments are defined according to the contents of the Wiley/NBS **Mass Spectral Database** and to fragmentation behavior. A search for structures with defined fragments (absence or presence of certain fragments) retrieves certain classes of compds. An application for checking a ref. spectrum is discussed. A search for structures similar to a target structure was developed for use in cases where a structure can be proposed for the **unknown** compd. The most closely related structures existing in the **database** will be selected, the resp. spectra often being the key for **interpretation** or **structure elucidation**, as illustrated by an example.

L22 ANSWER 150 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 118:80360 CA
TI Common **substructures** in **groups** of compounds exhibiting similar **mass spectra**
AU Scsibrany, H.; Varmuza, K.
CS Inst. Gen. Chem., Tech. Univ. Vienna, Vienna, A-1060, Austria
SO Fresenius' Journal of Analytical Chemistry (1992), 344(4-5), 220-2
AB Principal component projections of sets of **mass spectra** show clusters that contain compds. with common structural properties. The similarity of structures is investigated by an **automatic** search for large common **substructures** within the compds. of a cluster. Resulting spectra-structure-relations are helpful in **interpretation** of spectra.

L22 ANSWER 151 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 118:80205 CA
TI **Substructure identification** from **neutral loss information** of **mass spectra**
AU Hong, Qunfa; Zhu, Damo; Yang, Boyu; Xu, Chongde; Lu, Peizhang
CS Dalian Inst. Chem. Phys., Chin. Acad. Sci., Dalian, 116012, Peop. Rep. China
SO Fenxi Huaxue (1992), 20(10), 1117-20
LA Chinese
AB A **computer** program which is a functional part of ASES/MS **structure elucidation** system was developed for **substructure** identification from

neutral loss information. It is based on the **neutral loss mass spectra** of various functional **groups** and of one functional **group** in different structural environments, and **substructures** possibly contained in an **unknown** compd. are inferred from the primary and secondary **neutral loss information** of the **mass spectrum**.

L22 ANSWER 179 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 114:135337 CA
TI Exact mass **probability** based matching of high-resolution **unknown mass spectra**
AU Loh, Stanton Y.; McLafferty, Fred W.
CS Baker Chem. Lab., Cornell Univ., Ithaca, NY, 14853-1301, USA
SO Analytical Chemistry (1991), 63(6), 546-50
AB **Unknown mass spectra** measured with millimass accuracy can be matched (for quant. anal.) against a comprehensive unit-**mass**-resoln. **data base** of electron ionization spectra by utilizing its information on mol. elemental compns. and known correlations of common neutral species lost in ion dissocns. Adding this exact (E) mass capability to the **probability**-based matching (PBM) **algorithm** provides substantial performance improvements. Using matching criteria that retrieve 80% of the correct answers, EPBM increases the reliability of retrieving a spectrum of the same structure from 23% to 39%; accepting structural differences to which **mass spectrometry** is insensitive (class IV matches), EPBM increases the reliability from 44% to 71%, halving the no. of wrong answers. Similarly, for EPBM only 6% of best matches are incorrect (Class IV) vs. 10% by PBM.

L22 ANSWER 199 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 111:123146 CA
TI Development of **algorithms** for **automated** elucidation of **spectral** feature/**substructure** relationships in tandem **mass spectrometry**
AU Wade, A. P.; Palmer, P. T.; Hart, K. J.; Enke, C. G.
CS Dep. Chem., Michigan State Univ., East Lansing, MI, 48824, USA
SO Analytica Chimica Acta (1988), 215(1-2), 169-86
AB A **pattern-recognition** artificial-intelligence program, referred to as MAPS (method for analyzing patterns in **spectra**), is described for the **identification** of relationships that exist between the presence of **substructures** in mols. and the characteristic features they produce in **mass spectrometry** (MS) and tandem MS **data**. The MAPS **algorithm** discovers these relationships by intelligent anal. of a **data base** of MS and tandem MS spectra. The relationships found are expressed as rules, which may then be used to identify characterized **substructures** in **"unknowns"**. No prior knowledge of fragmentation pathways or rearrangements is assumed in the rule-generation process. While MAPS currently uses MS and tandem MS data, the approach (and much of the software) is equally suited to multiple-stage **mass spectrometric data**.

L22 ANSWER 208 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 110:91513 CA
TI **Computer** program for **post-translational** modification site assignment in proteins using fast atom bombardment **mass spectral data**
AU Pucci, P.; Sepe, C.

CS Dip. Chim. Org. Biol., Univ. Napoli, Naples, I-80134, Italy
SO Biomedical & Environmental Mass Spectrometry (1988), 17(4), 287-91
AB A **computer** program allowing **post-translational** modification sites assignment in proteins has been developed. The program has been constructed to elaborate **data** obtained from fast-atom-bombardment **mass spectrometric** mapping of polypeptides. The **mass values** of peptide(s) which cannot be assigned into the protein primary structure are elaborated by the program, which allows identification of the modified peptide(s) as well as the nature of the modifying **group**(s). This procedure has been applied to different kinds of **post-translational** events using three proteins as a model.

L22 ANSWER 210 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 110:22902 CA
TI ASES/MS: an **automatic structure elucidation** system for organic compounds using **mass spectrometric data**
AU Zhu, Damo; She, Jianwen; Hong, Qunfa; Liu, Renyu; Lu, Peichang; Wang, Luoqiu
CS Dalian Inst. Chem. Phys., Chin. Acad. Sci., Dalian, Peop. Rep. China
SO Analyst (Cambridge, United Kingdom) (1988), 113(8), 1261-5
AB A series program, which consists of a **library** search and an intelligence **interpretation** program, has been developed for compd. identification. The **library** search program uses a new combined forward and reverse search principle. The intelligence program assigns a mol. structure to an org. compd. by using spectrum-structure correlation rules based on about 25,000 ref. spectra.

L22 ANSWER 219 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 108:5274 CA
TI Multidimensional **computer** evaluation of **mass spectra**
AU Neudert, R.; Bremser, W.; Wagner, H.
CS BASF A.-G., Ludwigshafen, D-6700, Fed. Rep. Ger.
SO Organic Mass Spectrometry (1987), 22(6), 321-9
AB The generation of a **mass spectral interpretation** system is described that is usable both as part of a multidimensional system, and independently for the anal. of **mass spectra** only. The knowledge base is a structure-oriented **mass spectral data** collection consisting of some 42,000 spectra and topologies. The **comparison** of selected **mass spectral** properties such as similarity, **neutral losses**, and ion series of the **unknown** with the equiv. properties of the **library** spectra results in a set of corresponding structures. Subsequent **substructure** anal. yields a histogram of **substructure** frequencies contg. information about their statistical relevance. The relevant **substructure** set may be recombined to produce a structure proposal, as is demonstrated for 1-acetyl-2-methoxy-4-trimethylsilyloxybenzene. In a 2nd example, the relevant **substructures** derived by the **interpretation** system are used as input for the 13C-NMR **substructure** generator. This procedure reduces the soln. space of the structure prediction **algorithm** considerably. Besides the spectrum **interpretation**, addnl. possibilities are available. The **substructure** search enables, for example, a look for **mass spectrometric** reaction centers. Beyond that, **substructure** anal. is applicable to the detn. of structural features typical of certain combinations of **neutral**

losses and/or characteristic fragments.

L22 ANSWER 224 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 106:137624 CA
TI An **expert** system for organic structure determination
AU Curry, Bo
CS Chem. Syst. Dep., Hewlett-Packard Lab., Palo Alto, CA, 94304-1209, USA
SO ACS Symposium Series (1986), 306(Artif. Intell. Appl. Chem.), 350-64
AB An **expert** system which **interprets** low-resoln. **mass spectra**, IR **spectra**, and other user-supplied information and produces a list of functional **groups** present in an **unknown** org. compd. was described. The input data were **interpreted** as evidence supporting the presence or absence of each of the over 900 functional **groups** and org. **substructures** represented in the knowledge base. This evidence was then combined by an inference engine to det. the **probability** that the **group** was present. Each type of input spectra was **interpreted** by a sep. module, which had private internal data structures; these modules can use different techniques and even be written in different **computer** languages. The modular architecture was designed to allow new modules **interpreting** different types of spectra to be easily incorporated into the system. A major goal was the redn. of the no. of false pos. assertions.

L22 ANSWER 229 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 105:142772 CA
TI A spectral matching system for MS/MS data
AU Cross, K. P.; Enke, C. G.
CS Dep. Chem., Michigan State Univ., East Lansing, MI, 48824, USA
SO Computers & Chemistry (Oxford, United Kingdom) (1986), 10(3), 175-81
AB An **automated mass spectrometry/mass spectrometry** (MS/MS) **search** program was developed which allows the user to match an **unknown** MS/MS spectrum against either primary or secondary spectra in a ref. **data base**. The program employs several matching techniques for flexibility and avoids data compression or dependence on theor. spectral properties. The strategy of the program is to eliminate the majority of candidate MS/MS spectra by prefiltering the candidates through inverted data files. An intensity-based matching **algorithm** then dets. 7 match factors to completely characterize the correspondence between the **unknown** and each remaining candidate spectrum. Parabolic **fits** to quotient **spectra** are used, with limited success, to mask some deviations in spectra taken under different conditions. An expt. to characterize the program used 500 **mass spectra** from an old **data base** as **unknowns** for matching against the current MS/MS **data base**. The program retreived an identical or structurally closely related ref. compd. (when no identical compd. was present), 93% of the time.

L22 ANSWER 234 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 105:32368 CA
TI **Automation** of **structure elucidation** from **mass spectrometry-mass spectrometry data**
AU Cross, K. P.; Palmer, P. T.; Beckner, C. F.; Giordani, A. B.; Gregg, H. G.; Hoffman, P. A.; Enke, C. G.
CS Dep. Chem., Michigan State Univ., East Lansing, MI, 48824, USA
SO ACS Symposium Series (1986), 306(Artif. Intell. Appl. Chem.), 321-36

AB A system was designed to **automate** the extn. of structural **information** from **mass spectrometry-mass spectrometry** (MS/MS) **spectra**. Currently operational elements in this system include **data bases** for MS/MS spectra and mol. structures, spectrum matching programs, and a structure generator. Individual spectra within the complete set of MS/MS spectra are related to the mol. **substructures** from which they arise. The correlations between individual MS/MS spectra and specific **substructures** can be detd. by identifying the compds. that have matching MS/MS **spectra**, and then **identifying** the **substructures** they have in common. These correlations can supply identified **substructures** to a mol. structure generator such as GENOA. This empirical scheme assumes no knowledge of the fragmentation process, ion structures, or rearrangements.

L22 ANSWER 247 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 103:78840 CA
TI Retrieval and **interpretative computer** programs for **mass spectrometry**
AU McLafferty, Fred W.; Stauffer, Douglas B.
CS Chem. Dep., Cornell Univ., Ithaca, NY, 14853-1301, USA
SO Journal of Chemical Information and Computer Sciences (1985), 25(3), 245-52
AB Using the modern gas chromatograph/**mass spectrometer** (GC/MS), an **interpreter** may be faced with >100 **unknown** electron-ionization **mass spectra** per h. For this problem the **Probability** Based Matching (PBM) program yields real-time identifications for such a GC/MS output. Because GC sepn. can be incomplete, PBM employs reverse searching for improved identification of mixt. components. Forward searching, which is more specific for pure samples, is also **automatically** incorporated by matching the residual spectra obtained by subtracting the best matching ref. spectra from the **unknown**. The 81,000 different spectra of 68,000 different compds. of the Wiley/NBS **data base** were measured under a wide variety of exptl. conditions; to compensate for this variability, PBM employs peak "flagging" and abundance "caling". These and other values reflecting the degree of match are converted statistically into a single "reliability" value directly indicating the **probability** that the structure prediction is correct. With these improvements the 1st answer retrieved for pure and 60% mixt. components was correct, or difficult to distinguish from the correct answer by **mass spectrometry**, in 97% and 93% of cases, resp. If the **unknown** is not represented in the ref. file, the Cornell Self-Training **Interpretative** and Retrieval System predicts its mol.wt., no. of Cl and Br atoms, and substructural features present. For 589 **substructures**, a quant. "reliability" value is assigned to the STIRS prediction.

L22 ANSWER 254 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 102:142532 CA
TI **Computer**-aided **identification** of compounds by **comparison** of **mass spectra**
AU Domokos, L.; Henneberg, D.; Weimann, B.
CS Max-Planck Inst. Kohlenforsch., Muelheim/Ruhr, Fed. Rep. Ger.
SO Analytica Chimica Acta (1984), 165, 61-74
AB A new identity-oriented **search** procedure for **mass spectral libraries** (IDS) was developed by extending the similarity search system SISCOM. The aim of IDS is an exact identification of pure compds. and mixts. on

the basis of their **mass spectra**. The concepts and methods applied, e.g., filtering, feature selection, and optimization with **pattern recognition**, are described. Characteristics of IDS are summarized and demonstrated for several examples.

L22 ANSWER 270 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 99:63575 CA
TI Reproducibility as the basis of a similarity index for continuous variables in straightforward **library** search methods
AU Cleij, P.; Van't Klooster, H. A.; Van Houwelingen, J. C.
CS Anal. Chem., State Univ. Utrecht, Utrecht, 3522 AD, Neth.
SO Analytica Chimica Acta (1983), 150(1), 23-36
AB Straightforward **library** search methods, aiming at identification of (org.) compds. and based on **comparison** of anal. **data** for continuous variables, are considered with respect to a definition of the similarity of data. In the context used, the main object of such a search method is simply the retrieval of the ref. data of the **unknown** compd. The proposed similarity index has the form of a significance **probability** (P value), a quantity originating from the general theory of hypothesis testing ,and can be calcd. from a statistical model of the reproducibility of the quantities used for comparison. The index is defined in general terms, but is intended for applicability to **library** search methods for different types, or combinations, of anal. data. It is primarily designed for use in situations in which the application of very large **data bases** suffers from the generally low (interlab.) reproducibility of the data.

L22 ANSWER 294 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 91:221162 CA
TI A combined **forward-reverse library** search system for the **identification** of low-resolution **mass spectra**
AU Kwiatkowski, J.; Riepe, W.
CS Inst. Spektrochem. Angew. Spektrosk., Dortmund, D-4600/1, Fed. Rep. Ger.
SO Analytica Chimica Acta (1979), 112(3), 219-31
AB A combined **forward-reverse library search** routine for low-resoln. **mass spectra** is described. The routine requires binary-coded spectra. Masses and **peak intensities** are used for **spectral comparison**. On the basis of 3 possible search strategies, this routine is adaptable to anal. problems. The program was tested for 25,000 spectra from the ISAS, MSDC and EPA **mass spectra libraries**. The program is written completely in FORTRAN IV.

L22 ANSWER 296 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 91:101576 CA
TI Identification of components in mixtures by a mathematical analysis of **mass spectral data**
AU Rasmussen, G. T.; Hohne, B. A.; Wieboldt, R. C.; Isenhour, T. L.
CS Dep. Chem., Univ. North Carolina, Chapel Hill, NC, 27514, USA
SO Analytica Chimica Acta (1979), 112(2), 151-64
AB Math. techniques for the identification of components in mixts. from the **mass spectra** of a series of related mixts. are described. The approach is analogous to **library search** methods in that **spectra** from a ref. collection are **compared** with a multidimensional **unknown**. Searches are

conducted with a **library** file contg. approx. 17,000 **mass spectra**. Results for the analyses of several mixts. are reported, to illustrate the effectiveness of the method.

L22 ANSWER 306 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 90:114533 CA
TI SISCOM - a new **library search** system for **mass spectra**
AU Damen, H.; Henneberg, D.; Weimann, B.
CS Max-Planck-Inst. Kohlenforsch., Muelheim/Ruhr, Fed. Rep. Ger.
SO Analytica Chimica Acta (1978), 103(4), 289-302
AB SISCOM is a **library search** system for **mass spectrometry** which is based on a new method of coding **spectra** by selecting the most important **peaks** within homologous ion series, and on a multiple factor assessment of the result. Examples demonstrate the ability of the system to identify various compds., even from mixts. or by ref. spectra which differ from those measured. SISCOM is esp. suitable for detecting structural similarities like common **substructures**, even in cases where no similarity can be recognized by visual comparison of patterns or by human **interpretation** of the spectrum.

L22 ANSWER 341 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 82:67893 CA
TI **Automated identification** of **mass spectra** by the reverse **search**
AU Abramson, Fred P.
CS Sch. Med., George Washington Univ., Washington, DC, USA
SO Analytical Chemistry (1975), 47(1), 45-9
AB A new method for the **automatic identification** of **mass spectra** which used the **library spectrum** as the basis of the **comparison** is described. This process, called reverse search, is contrasted with other methods for **mass spectral library searches** where the **unknown spectrum** itself becomes the basis. The reverse search is shown to be fully **automated**, requiring no operator judgment to output qual. and quant. data. The other significant feature of a reverse search is its inherent rejection of interference. A specific compd. obscured by other compds. can still be identified by this method. A no. of areas of routine anal. are suggested where this system could have significant application. The **automated** identification process is esp. valuable when operating a gas chromatograph-**mass spectrometer** system.

L22 ANSWER 368 OF 387 CA COPYRIGHT 2005 ACS on STN
AN 75:64256 CA
TI Submolecular **group** analyses of high resolution **mass spectra** of peptides and other sequence molecules
AU Kunderd, A.; Spencer, R. B.; Budde, W. L.
CS Mass Spectrom. Cent., Purdue Univ., Lafayette, IN, USA
SO Analytical Chemistry (1971), 43(8), 1086-90
AB Submol. **group** masses are defined as the sums of exact masses of **groups** of atoms which form **substructures** of moles. A std. type of **computer** program was utilized to search rapidly and thoroughly for combinations of submol. **groups** whose calcd. masses were within a few millimass units of measured **masses** from high resolution **mass spectra**. The advantages of these combinations as a means of quickly identifying key information contg. ions are evaluated. The submol. **group** approach is of value and

the approach is applied to the **mass spectra** of a simple org. mol. and a tetrapeptide.

=> log y
STN INTERNATIONAL LOGOFF AT 17:10:55 ON 10 MAY 2005